AUTHORS' RESPONSE TO THE COMMENTS OF THE EDITOR

Dear Tomáš Uxa and co-authors,

as you have certainly noticed two reports from the second review round are online. While one of them mainly requests to include permafrost sites in forest areas, the other report is very critical. It raises four major comments - all valid and constructive in my opinion. First of all, your MAPT and ALT 'observations' seem to be generated by linear interpolation. This appears to heavily undermine the validation of your model. Second, reviewer #3 points you to a valuable database for circum-arctic active layer monitoring. Third, the reviewer requests you to assess the performance of a baseline model to better understand the benefit of your approach. Last, you are pointed to two other approaches (Nicholas, 1995; Romanovsky, 1996) that also indirectly infer bulk material properties of permafrost similar to your method. This calls for a moderation when presenting your method.

In summary, I consider the comments on the validation and the baseline model as crucial for continuing to consider your article for publication in TC. Despite the reviewer recommendation, I invite you to submit a revised manuscript. If you should see yourself incapable of addressing these comments, please consider to re-submit your manuscript as a shorter or more technical report.

Best,
The editor, Johannes Fürst

Dear Editor,

On behalf of co-authors, I am submitting the second revision of the manuscript ID EGUSPHERE-2024-2989 entitled "Simple analytical–statistical models (ASMs) for mean annual permafrost table temperature and active-layer thickness estimates" by Tomáš Uxa, Filip Hrbáček, and Michaela Kňažková.

We responded point-by-point to all the reviewers' comments and suggestions and made corresponding revisions in the manuscript. Our responses are included at the end of this text, starting on page 2. Please note that the reviewers' comments are prefixed by bold RC (1 or 3), while the authors' comments are prefixed by bold AC. All the reviewers' comments are also in blue in order to distinguish them from the authors' responses.

We included additional 11 forest and 1 shrub site so that their proportion of the total dataset (now counting 55 sites) has substantially increased in the revised manuscript. Correspondingly, we also recalculated the results and revised the text, tables and figures. However, the results changed to minor extents, which suggests that the analysis was already robust enough.

The referee #3 mostly criticizes our approach, but we lack any kind of constructive suggestions and specific recommendations on how we should improve the manuscript. Therefore, in our reply, we respond to the particular points of the criticism and unsubstantiated statements that mostly concern the validation of the proposed models. However, we performed the validation using standard approaches for MAPT/ALT determination, which are widely accepted by the research community. If this approach was considered illegitimate or seriously flawed, it would call into question the majority of results published to date on the active-layer thermal regime. The major points of our response to the referee #3 (mostly related to the MAPT and ALT for validation) were also incorporated into the revised manuscript, especially Sect. 3 "Model evaluation" (please see our detailed response).

We believe that the improvements made in the revised manuscript are sufficient.

Thank you very much for considering the revised manuscript.

Yours sincerely,
Tomáš Uxa

AUTHORS' RESPONSE TO THE COMMENTS OF THE REFEREE #1

RC1: The manuscript has been significantly improved by the revision work. The presentation and discussion of the proposed methodology is clearer, and the validation is more convincing. Meanwhile I have a few minor concerns regarding the discussion of the underlying assumptions and the chosen set of validation sites, see my comments below. Thus I recommend a minor revision of the manuscript prior to its publication in TC.

AC: Thank you for your review.

RC1: l 255-256: "Although ASMs utilize only thawing and freezing indices from two depth levels within the active layer as inputs, they inherently account for the natural variability of ground physical properties in the intermediate layer between these two depths." I understand better now what the authors mean here, nevertheless I think that it should be highlighted that with the developed methodology the variability of ground physical properties is handled by means of an averaging operation, resulting in a temporally (seasonal) and spatially (within all the layer between z1 and z2) average of kt/kf.

AC: We revised the sentence as follows: "Although ASMs utilize only thawing and freezing indices from two depth levels within the active layer as inputs, they inherently account for the natural variability of ground physical properties in the intermediate layer between these two depths that is expressed in terms of annual and seasonal means of the thermal conductivity ratio and edaphic term, respectively."

RC1: l 269-270: "Of course, ASMs in principle also treat them as constants, but their values are representative for individual years (Eq. 8) or thawing seasons (Eq. 27)" Here the sources of temporal variability (e.g.: temporal variation of water content) should be explicated.

AC: We revised the sentence as follows: "Of course, ASMs also treat them as constants, but their values are annual or seasonal means that reflect the variations in ground physical properties over time mainly due to changes in water content and as such they are representative for individual years (Eq. 8) or thawing seasons (Eq. 27). This is a major improvement over other analytical or statistical models…"

RC1: Table C1 – the increase in number and diversity of the considered validation sites is a strong improvement of the manuscript. Nevertheless the proportion of forested sites is small (5/43, ~12%) compared to the proportion of the permafrost area covered by boreal forest (55%, Stuenzi et al., 2021). Additional sites in boreal forest areas should be added, or at least the possible biases linked to the used sampling of permafrost areas conditions should be discussed.

Stuenzi, S. M., Boike, J., Gädeke, A., Herzschuh, U., Kruse, S., Pestryakova, L. A., Westermann, S., and Langer, M.: Sensitivity of ecosystem-protected permafrost under changing boreal forest structures, Environ. Res. Lett., 16, 084045, https://doi.org/10.1088/1748-9326/ac153d, 2021.

AC: We included additional 11 forest and 1 shrub site so that their proportion of the total dataset (now counting 55 sites) has substantially increased in the revised manuscript. However, note that the results for shrub and forest sites as well as the total ones changed to minor extents, which suggests that the analysis was already robust enough.

AUTHORS' RESPONSE TO THE COMMENTS OF THE REFEREE #3

RC3: In this paper, the authors propose set of semi-analytical models for estimating active layer thickness (ALT) and mean annual permafrost temperatures (MAPT) in permafrost-affected soils using only ground temperatures measured in the active layer. The presented equations are relatively straightforward algebraic derivations from well known equilibrium models of permafrost temperature (TTOP) and thaw (Stefan model). The authors validate their approach using ground temperature data from 43 sites in the Arctic spanning a wide range of surface conditions.

AC: Thank you for your review, even though it mostly criticizes our approach. Unfortunately, we lack any kind of constructive suggestions and specific recommendations on how we should improve the manuscript. Therefore, in our reply, we aim to provide you with a thorough response to the particular points of your criticism and unsubstantiated statements. The major points of our response (mostly related to the MAPT and ALT for validation) were also incorporated into the revised manuscript, especially Sect. 3 ”Model evaluation”.

Please note that our validation dataset covers a much wider area than the Arctic and a broader range of permafrost environments. In addition to the Arctic, it includes other Earth’s major permafrost regions such as Antarctica, Qinghai-Tibetan Plateau and European Alps (now counting 55 sites), where high-quality ground temperature measurements are available. Although this number may seem low, it reflects the fact that the sites were carefully selected with respect to data availability, overall continuity of datasets, and sensor spacing.

RC3: It is unclear to me whether or not the equations presented by the authors in this work are truly novel. I have not seen them before in this particular form anywhere in the literature, so I will give them the benefit of the doubt and assume that they are. However, the derivations are relatively straightforward and follow from long established equilibrium models, so I think that the added value is relatively low, all other assumptions and limitations of such equilibrium models notwithstanding.

AC: It is true that the analytical-statistical models for MAPT and ALT given by Eq. (8) and (27) were derived from long-established models (TTOP and Stefan models). However, to our knowledge, Eq. (8) and (27) themselves have never been published so far anywhere (papers, monographs, or conference proceedings) and thus they are totally novel. We also believe that if any derivation is simple and straightforward, this is an advantage rather than the opposite. No matter what it is based on, if it is new.

The added value of our solutions is a matter of largely subjective opinion, but we respectfully disagree with the statement that “all other assumptions and limitations” of the novel models are the same as those of the TTOP and Stefan models. Our solutions address the issue of the utilization of the ground physical properties for the modelling. They reflect the variations in ground physical properties over time mainly due to changes in water content and as such they are representative for individual years (Eq. 8) or thawing seasons (Eq. 27). This is a major improvement over other analytical or statistical models.

RC3: The equations themselves are indeed perhaps useful in some problem settings and would certainly fit well in a technical report. However, I do not think that the level of novelty or scientific rigor of the validation meets the editorial standard for a full journal article.

From the scientific perspective, this article has two major flaws:

1. The “observed” MAPT and ALT are not actually observations, but are rather obtained by linear interpolation of observed ground temperatures in both the active layer and permafrost via linear interpolation. As far as I could tell, the authors do not report the depths of the temperature sensors which were used to calculate these pseudo-observations, which is also problematic. Such approximations could potentially be acceptable if the interpolation is over a short distance (<10 cm or so) but otherwise cannot be expected to realistically determine the actual MAPT and ALT.

AC: Thank you for your opinion, however, we disagree that we are generating “pseudo-observations”. The interpolation of ALT is widely used and accepted and in important permafrost environments such as mountains or high polar regions (extensive occurrence of exposed bedrock surfaces or gravelly to blocky materials) it is the only method for ALT estimates because manual probing is impossible. Further, if we are seeking the position/depth of 0 °C isotherm in the particular borehole/profile (site) there is no other way than using and interpolating data from ground temperature measurements as these cannot be spatially continuous. By contrast, there is always some sensor spacing that tends to increase with depth and interpolating between them is therefore a standard practice (see e.g. Measurements Recommendations and Guidelines at Global Terrestrial Network for Permafrost portal at http://gtnpdatabase.org).

To our knowledge, there is no standard so far, which recommends particular interpolation method or spacing between thermometers including the justification why this particular method is better than others. However, Measurements Recommendations and Guidelines at Global Terrestrial Network for Permafrost (at http://gtnpdatabase.org) suggest linear interpolation and sensor spacing within the active layer of 0.2, 0.4, 0.8, 1.2, 1.6, 2, 2.5 m). The contribution by Riseborough (2008) deals with the possible inaccuracies of the interpolation, but unfortunately these ideas were not developed further, remaining this issue unresolved. It is also necessary to state that the assumption of Riseborough (2008) is solely based on modelling data and have not been verified in any way in the field. Yet, considering this general knowledge, the vast majority of localities was selected so that the maximum distance between sensors was lower than 25 cm and 50 cm (for ALT <100 and >100 cm, respectively; please note that this is much finer depth resolution/smaller sensor spacing than recommended – see above), which on the one hand significantly reduced the number of datasets available for the validation (sensor spacing is greater at numerous sites), but on the other hand increased the accuracy of the observed MAPT and ALT. However, several exceptions were made to this rule in order to balance the different surface covers in the validation dataset.

Regarding the use of the interpolated temperature to estimate MAPT, the error is much lower in comparison to ALT. Usually, there is a low gradient in the mean annual ground temperature near the base of the active layer/permafrost table and so the temperature difference observed at the deepest sensor in the active layer and at the topmost sensor in permafrost was lower than 0.2 °C. In other words, if we would use measured data from the first (=uppermost) sensor located in permafrost only the overall effect on the accuracy of the MAPT estimate would be negligible.

We quantified and included in the revised manuscript that the temperature differences between the observed MAPT and the temperature of the closest sensor used for the interpolation differed by less than 0.1 °C in ~65 % of cases and by less than 0.2 °C in ~90 %. The distance between the observed ALT and the closest temperature sensor used of the interpolation was less than 10 cm in ~80 % of cases and less than 20 cm in almost 100 %. This gives the maximum possible deviations of observed MAPT and ALT from their actual values. We therefore think that the accuracy should be mostly acceptable also based on the reviewer's opinion (=when interpolation is over a short distance [<10 cm or so]).

RC3: This also likely explains why the model appears to perform so well at many of the sites. According to the authors' own description, their method effectively amounts to a form of linear extrapolation from the two measured active layer temperatures. By generating pseudo-observations via linear interpolation, the authors are effectively generating data which, by construction, will be well predicted by such a linear extrapolation, in most cases. It is unclear why the authors did not attempt to use actual active layer measurements such as those from the Circum-arctic Active Layer Monitoring (CALM) network.

AC: The utilization of data from CALM sites is not feasible for this kind of validation from following reasons:

1) Please note that the MAPT and ALT estimates using Eq. (8) and (27) were exclusively based on the depth pairs from within the active layer and especially its uppermost part in most instances. By contrast, the observed MAPT and ALT was determined using the linear interpolation between the deepest sensor in the active layer and at the topmost sensor in permafrost. This means that the sensors used to determine the observed and modelled values differed in all instances! There was not a single identical pair of sensors used for both validation and modelling so your assumption is incorrect.

2) The data presented in the CALM database are typically mean values from (usually) 121 measurements within grids of 100 x 100 m, which typically show a high spatial variability within the grid, as was documented in studies from different Arctic as well as Antarctic regions. Even if we would have ground temperature data from

boreholes at the individual CALM sites, we could not validate active-layer thickness determined from borehole data against mean thaw depth from the grid of 100 x 100 m.

3) Manual thaw-depth probing measures a physical state of the ground using rigid rods that are pushed vertically into the ground to the depth at which ice-bonded material provides firm resistance. Since ice formation at sub-zero temperatures is complex and ground freezing characteristic curves differ between substrates, manually determined thaw depth may not necessarily correspond to the position of 0 °C isotherm, which is obtained by models.

4) Manual thaw-depth probing is impossible in mountains or high polar regions with extensive occurrence of exposed bedrock surfaces or gravelly to blocky materials. Consequently, there are no manual thaw-depth measurements from such sites.

5) Numerous CALM sites, such as those from Svalbard, Switzerland or Mongolia, therefore only utilize ground temperature measurements at depths of 0, 1, 2, 3 m and their subsequent interpolation to determine ALT (see https://www2.gwu.edu/~calm/data/north.htm). Hence, a substantial part of the CALM database, which you recommend to use for model validations, also generates "pseudo-observations" and should be considered unreliable according to your opinion.

6) We think that using and interpolating ground temperature measurements makes the validation dataset homogeneous because uniform methods can be used across all sites. This would be impossible for CALM sites where thaw-depth is measured manually by different persons and frequently only once a year, which means that the maximum thaw depth (=active-layer thickness) is in reality rarely reported from these sites. This is unsuitable for validations of our models.

RC3: 2. Although the validation is ostensibly extensive in terms of the number of field sites, the authors provide no baselines for comparisons. This is especially surprising considering that the authors' method is derived from the TTOP and Stefan models, both of which would serve as natural baselines. The authors might respond that the novelty of their method is in the fact that they do not require estimates of the frozen vs. thawed thermal conductivities, information which they do not have for many or all of the sites.

AC: Indeed, this is the major reason why we cannot use the TTOP and Stefan models themselves for the validations. To our knowledge, the data on thermal properties, but also on ground water content, are simply unavailable for the validation sites, which makes direct validation against the two models impossible. Further, using TTOP/Stefan models as reference would provide us with the accuracy of ASMs against these two models but not against observed data.

RC3: However, this would seem to ignore the fact that methods such as that of Hinkel and Nicholas (1995) and Romanovsky (1996), both of which are cited within, get around this simply by treating lumping these unknown terms into a linear parameter which can be estimated directly from the data. It would also be natural for the authors to compare their method to the similar and widely used approach of extrapolating the MAGT at the two measured points down to the zero-degree isotherm, which they themselves state is closely related.

AC: Please note that the edaphic term does not fully solve the issue of missing parameter. Since it combines the information on thermal and moisture properties, it also shows natural year-to-year variations caused by the variability of these ground properties. This means that it in principle leads to random over/underestimation of ALT over time, while our procedure considers the actual (average) state of the active layer in individual years or thawing seasons. Notably, the edaphic term is impossible to use at sites where you lack any data on ALT, which are crucial for the calculation of the edaphic term. This is one of the issues which we suggest to solve using ASMs.

RC3: In summary, although the equations derived by the authors may be of some interest to polar researchers and practitioners in the form of a technical report, the scientific novelty is modest at best, and the rigor of the validation is seriously lacking. Given that the paper has already gone through a full round of revision, I would recommend rejection.

AC: Your main criticism concerns the validation of the proposed ASMs. However, we performed the validation using standard approaches for MAPT/ALT determination, which are widely accepted by the research community. If this approach was considered illegitimate or seriously flawed, it would call into question the majority of results published to date on the active-layer thermal regime. We believe that we addressed and explained these issues sufficiently.